

JC658 U.S. PRO  
09/723512  
11/28/00



日本国特許庁  
PATENT OFFICE  
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されて  
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed  
with this Office.

出願年月日  
Date of Application:

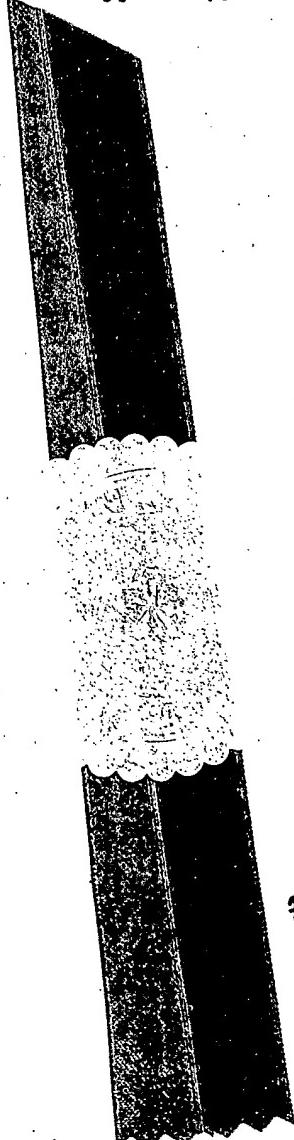
1999年11月30日

出願番号  
Application Number:

平成11年特許願第340468号

出願人  
Applicant(s):

ソニー株式会社

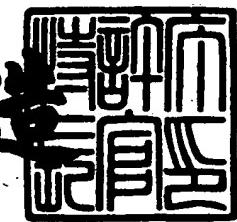


CERTIFIED COPY OF  
PRIORITY DOCUMENT

2000年 9月 8日

特許庁長官  
Commissioner,  
Patent Office

及川耕



出証番号 出証特2000-3072724

【書類名】 特許願  
【整理番号】 9900892203  
【提出日】 平成11年11月30日  
【あて先】 特許庁長官殿  
【国際特許分類】 G10L 3/00  
【発明者】  
【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社  
内  
【氏名】 南野 活樹  
【特許出願人】  
【識別番号】 000002185  
【氏名又は名称】 ソニー株式会社  
【代表者】 出井 伸之  
【代理人】  
【識別番号】 100082131  
【弁理士】  
【氏名又は名称】 稲本 義雄  
【電話番号】 03-3369-6479  
【手数料の表示】  
【予納台帳番号】 032089  
【納付金額】 21,000円  
【提出物件の目録】  
【物件名】 明細書 1  
【物件名】 図面 1  
【物件名】 要約書 1  
【包括委任状番号】 9708842  
【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識装置および音声認識方法、並びに記録媒体

【特許請求の範囲】

【請求項1】 ロボットに内蔵される音声認識装置であって、  
音声を認識する音声認識手段と、  
前記ロボットの状態に基づいて、前記音声認識手段を制御する制御手段と  
を備えることを特徴とする音声認識装置。

【請求項2】 制御手段は、前記ロボットの成長、感情、または本能の状態  
に基づいて、前記音声認識手段を制御する  
ことを特徴とする請求項1に記載の音声認識装置。

【請求項3】 前記制御手段は、前記ロボットの状態に基づいて、前記音声  
認識手段の音声認識精度を変更する  
ことを特徴とする請求項1に記載の音声認識装置。

【請求項4】 前記音声認識手段は、音声認識の対象とする語彙が登録され  
た辞書を記憶する辞書記憶手段を有し、  
前記制御手段は、前記辞書に登録された語彙に対して、前記ロボットの状態に  
対応する重みを付して音声認識を行うように、前記音声認識手段を制御する  
ことを特徴とする請求項1に記載の音声認識装置。

【請求項5】 前記音声認識手段は、音声認識の対象とする語彙が分けて登  
録された複数の辞書を記憶する辞書記憶手段を有し、  
前記制御手段は、前記複数の辞書に対して、前記ロボットの状態に対応する重  
みを付して音声認識を行うように、前記音声認識手段を制御する  
ことを特徴とする請求項1に記載の音声認識装置。

【請求項6】 前記音声認識手段は、音声認識の対象とする語彙が、他の語  
彙と対応付けられて登録された辞書を記憶する辞書記憶手段を有し、  
前記制御手段は、前記ロボットの状態に基づいて、前記辞書において、音声認  
識結果としての語彙に対応付けられている前記他の語彙を、最終的な音声認識結  
果として出力するように、前記音声認識手段を制御する  
ことを特徴とする請求項1に記載の音声認識装置。

【請求項7】 前記辞書には、音声認識の対象とする語彙が、その語彙と音響的に類似しているまたは意味的に関連する他の語彙と対応付けられて登録されている

ことを特徴とする請求項6に記載の音声認識装置。

【請求項8】 前記音声認識手段は、音声認識の対象とする語彙が登録される辞書を記憶する辞書記憶手段を有し、

前記制御手段は、前記ロボットの状態に基づいて、前記辞書に登録可能な語彙の数を制御する

ことを特徴とする請求項1に記載の音声認識装置。

【請求項9】 前記ロボットは、前記音声認識手段が出力する音声認識結果に基づいて、所定の行動を起こす

ことを特徴とする請求項1に記載の音声認識装置。

【請求項10】 ロボットに内蔵される音声認識装置の音声認識方法であつて、

音声を認識する音声認識ステップと、

前記ロボットの状態に基づいて、前記音声認識ステップを制御する制御ステップと

を備えることを特徴とする音声認識方法。

【請求項11】 ロボットに音声認識を行わせるために、コンピュータが実行するプログラムが記録されている記録媒体であつて、

音声を認識する音声認識ステップと、

前記ロボットの状態に基づいて、前記音声認識ステップを制御する制御ステップと

を備えるプログラムが記録されている

ことを特徴とする記録媒体。

#### 【発明の詳細な説明】

##### 【0001】

##### 【発明の属する技術分野】

本発明は、音声認識装置および音声認識方法、並びに記録媒体に関し、音声認

識機能を有するロボットに用いて好適な音声認識装置および音声認識方法、並びに記録媒体に関する。

#### 【0002】

##### 【従来の技術】

例えば、音響モデル、辞書、文法などに基づいて、音響スコアおよび言語スコアを求め、入力音声の音声認識結果としての単語系列が決定される音声認識技術が知られている。また、ある命令信号に対応して、ロボットの動作を制御する制御方法が、FA (Factory Automation) などのロボットに用いられている。さらに、これらの音声認識技術およびロボット制御技術を組み合わせ、音声認識結果に基づいて、ロボットの動作を制御するシステムも実現されている。そのようなシステムとしては、例えば、荷物の仕分けにおいて、音声認識を利用して荷物の振り分け先を変更するシステムがある。なお、ロボットの他、コンピュータのコントロール（コマンド入力等）などを、音声認識を利用して行うことも、近年広く行われるようになってきている。

#### 【0003】

また、上述のようないわゆる産業用ロボットの他、疑似ペットとしての、エンタテイメント用のロボット（ぬいぐるみ状のものを含む）においても、使用者の発話を音声認識し、その音声認識結果に基づいて、各種の行動を起こすものが製品化されつつある。

#### 【0004】

##### 【発明が解決しようとする課題】

ところで、特に、エンタテイメント用のロボットについては、実際のペットが成長していくのと同様の感覚を、使用者に享受させることが望まれる。即ち、例えば、最初は、使用者の発話を理解（音声認識）できないが、成長するにつれて、使用者の発話を徐々に理解することができるようすれば、ロボットの娛樂性を格段的に向上させることができる。

#### 【0005】

本発明は、このような状況に鑑みてなされたものであり、ロボットの成長等の状態にあわせて変化する音声認識処理を行うことができるようにするものである

【0006】

【課題を解決するための手段】

本発明の音声認識装置は、音声を認識する音声認識手段と、ロボットの状態に基づいて、音声認識手段を制御する制御手段とを備えることを特徴とする。

【0007】

制御手段には、ロボットの成長、感情、または本能の状態に基づいて、音声認識手段を制御させることができる。

【0008】

また、制御手段には、ロボットの状態に基づいて、音声認識手段の音声認識精度を変更させることができる。

【0009】

音声認識手段には、音声認識の対象とする語彙が登録された辞書を記憶する辞書記憶手段を設け、制御手段には、辞書に登録された語彙に対して、ロボットの状態に対応する重みを付して音声認識を行うように、音声認識手段を制御させることができる。

【0010】

また、音声認識手段には、音声認識の対象とする語彙が分けて登録された複数の辞書を記憶する辞書記憶手段を設け、制御手段には、複数の辞書に対して、ロボットの状態に対応する重みを付して音声認識を行うように、音声認識手段を制御させることができる。

【0011】

さらに、音声認識手段には、音声認識の対象とする語彙が、他の語彙と対応付けられて登録された辞書を記憶する辞書記憶手段を設け、制御手段には、ロボットの状態に基づいて、辞書において、音声認識結果としての語彙に対応付けられている他の語彙を、最終的な音声認識結果として出力するように、音声認識手段を制御させることができる。

【0012】

辞書には、音声認識の対象とする語彙を、その語彙と音響的に類似しているま

たは意味的に関連する他の語彙と対応付けて登録しておくことができる。

【0013】

音声認識手段には、音声認識の対象とする語彙が登録される辞書を記憶する辞書記憶手段を設け、制御手段には、ロボットの状態に基づいて、辞書に登録可能な語彙の数を制御させることができる。

【0014】

ロボットは、音声認識手段が出力する音声認識結果に基づいて、所定の行動を起こすものとすることができます。

【0015】

本発明の音声認識方法は、音声を認識する音声認識ステップと、ロボットの状態に基づいて、音声認識ステップを制御する制御ステップとを備えることを特徴とする。

【0016】

本発明の記録媒体は、音声を認識する音声認識ステップと、ロボットの状態に基づいて、音声認識ステップを制御する制御ステップとを備えるプログラムが記録されていることを特徴とする。

【0017】

本発明の音声認識装置および音声認識方法、並びに記録媒体においては、ロボットの状態に基づいて、音声認識処理が制御される。

【0018】

【発明の実施の形態】

図1は、本発明を適用したロボットの一実施の形態の外観構成例を示しており、図2は、その電気的構成例を示している。

【0019】

本実施の形態では、ロボットは、犬形状のものとされており、胴体部ユニット2には、その前端に頭部ユニット3が配設され、また、その前後左右の各隅部には、それぞれ、大腿部ユニット4A, 4B, 4C, 4Dおよび脛部ユニット5A, 5B, 5C, 5Dからなる脚部ユニット6A, 6B, 6C, 6Dが取り付けられている。さらに、胴体部ユニット2の後端には、尻尾ユニット1が取り付けられている。

れている。

### 【0020】

尻尾ユニット1と胴体部ユニット2、頭部ユニット3と胴体部ユニット2、大腿部ユニット4A乃至4Dそれぞれと胴体部ユニット2、および大腿部ユニット4A乃至4Dそれぞれと脛部ユニット5A乃至5Dそれぞれを連結する各関節機構には、アクチュエータとしてのモータ $7_1, 7_2, \dots, 7_N$ が配設されており、これにより、対応するモータ $7_1$ 乃至 $7_N$ を駆動することによって、尻尾ユニット1および頭部ユニット3を、x軸、y軸、z軸の3軸それぞれを中心とする方向に自在に回転させ、腿部ユニット4A乃至4Dを、x軸、y軸の2軸それぞれを中心とする方向に自在に回転させ、かつ脛部ユニット5A乃至5Dを、x軸の1軸を中心とする方向に回転させ得るようになっており、これにより、ロボットは、各種の行動を行うことができるようになっている。

### 【0021】

頭部ユニット3には、カメラ（CCD（Charge Coupled Device）カメラ）8、マイク（マイクロフォン）9、および圧力センサ10がそれぞれ所定位置に配設されており、また、胴体部ユニット2は、制御部11を内蔵している。カメラ8では、使用者を含む周囲の状況の画像が撮像され、マイク9では、使用者の音声を含む周囲の音声が集音される。また、圧力センサ10では、使用者等によって与えられる圧力が検出される。そして、カメラ8により撮像された周囲の状況の画像や、マイク9により集音された周囲の音声、圧力センサ10により検出された、使用者により頭部に与えられた圧力は、それぞれ画像データ、音声データ、圧力検出データとして、それぞれ制御部11に与えられる。

### 【0022】

各関節機構に対応する各モータ $7_1$ 乃至 $7_N$ については、それぞれに対応させてロータリエンコーダ $12_1$ 乃至 $12_N$ が設けられており、各ロータリエンコーダ $12_1$ 乃至 $12_N$ では、対応するモータ $7_1$ 乃至 $7_N$ の回転軸の回転角度が検出される。ロータリエンコーダ $12_1$ 乃至 $12_N$ で検出された回転角度は、それぞれ角度検出データとして制御部11に与えられる。

### 【0023】

制御部11は、カメラ8からの画像データ、マイク9からの音声データ、圧力センサ10からの圧力検出データ、およびロータリエンコーダ $12_1$ 乃至 $12_N$ それぞれからの角度検出データに基づいて、周囲の状況や自分の姿勢等を判断するとともに、予めインストールされている制御プログラムに基づいて続く行動を決定し、その決定結果に基づいて、必要なモータ $7_1$ 乃至 $7_N$ を駆動させるようになっている。

#### 【0024】

これにより、ロボットは、尻尾ユニット1や、頭部ユニット2、各脚部ユニット $6A$ 乃至 $6D$ を動かして所望状態にし、自律的に行動する。

#### 【0025】

次に、図3は、図2の制御部11の構成例を示している。

#### 【0026】

制御部11は、CPU (Central Processing Unit) 20、プログラムメモリ21、RAM (Random Access Memory) 22、不揮発性メモリ23、インターフェース回路 (I/F) 24、およびモータドライバ25が、バス26を介して接続されて構成されている。

#### 【0027】

CPU (Central Processing Unit) 20は、プログラムメモリ21に記憶されている制御プログラムを実行することにより、ロボットの行動を制御する。プログラムメモリ21は、例えば、EEPROM (Electrically Erasable Programable Read Only Memory) 等で構成され、CPU 20が実行する制御プログラムや必要なデータを記憶している。RAM 22は、CPU 20の動作上必要なデータ等を一時記憶するようになっている。不揮発性メモリ23は、後述するような感情／本能モデルや、行動モデル、成長モデル、単語辞書、音韻モデル、文法規則等の、電源がオフ状態とされた後も保持する必要のあるデータを記憶するようになっている。インターフェース回路24は、カメラ8や、マイク9、圧力センサ10、ロータリエンコーダ $12_1$ 乃至 $12_N$ から供給されるデータを受信し、CPU 20に供給するようになっている。モータドライバ25は、CPU 20の制御にしたがい、モータ $7_1$ 乃至 $7_N$ を駆動（ドライブ）するための駆動信号を、モー

タ $T_1$ 乃至 $T_N$ に供給するようになっている。

#### 【0028】

制御部11は、CPU20において、プログラムメモリ21に記憶された制御プログラムが実行されることにより、機能的に、図4に示すような構成となって、ロボットの行動を制御する。

#### 【0029】

即ち、図4は、制御部11の機能的な構成例を示している。

#### 【0030】

センサ入力処理部30は、外部から与えられる刺激や、外部の状態を、マイク9や、カメラ8、圧力センサの出力に基づいて認識し、成長モデル部31、感情／本能モデル部32、および行動決定部33に供給するようになっている。なお、センサ入力処理部30は、必要に応じて、成長モデル部31の出力を用いて処理を行うようになっている。この点については後述する。

#### 【0031】

成長モデル部31は、ロボットの成長の状態を表す成長モデルを管理しており、センサ入力処理部30の出力や、時間経過に基づいて、ロボットの成長状態を変化させ、その変化後の成長状態を、センサ入力処理部30および行動決定部33に供給するようになっている。

#### 【0032】

感情／本能モデル部32は、ロボットの感情と本能の状態をそれぞれ表現する感情モデルと本能モデルを管理しており、センサ入力処理部30の出力や、行動決定部33の出力、時間経過に基づいて、感情モデルや本能モデルを規定するパラメータを変更することで、ロボットの感情や本能の状態を変化させるようになっている。この変化後の感情や本能の状態は、行動決定部33に供給されるようになっている。

#### 【0033】

行動決定部33は、行動モデル記憶部33Aを内蔵しており、この記憶内容や、センサ入力処理部30の出力、成長モデル部31で管理されている成長モデル、感情／本能モデル部32で管理されている感情モデルおよび本能モデル、時間

経過に基づいて、その後にロボットが行う行動を決定して、その情報（以下、適宜、行動情報という）を、姿勢遷移部34に供給するようになっている。

#### 【0034】

姿勢遷移部34は、行動決定部33から供給される行動情報に対応する行動をロボットに行わせるためのモータ $7_1$ 乃至 $7_N$ の回転角や回転速度等の制御データを演算し、モータ制御部35に出力する。

#### 【0035】

モータ制御部35は、姿勢遷移部34からの制御データにしたがって、モータ $7_1$ 乃至 $7_N$ を回転駆動するようになっている。

#### 【0036】

以上のように構成される制御部11では、入力センサ処理部30において、カメラ8から供給される画像データや、マイク9から供給される音声データ、圧力センサ10から出力される圧力検出データに基づいて、特定の外部状態や、使用者からの特定の働きかけ、使用者からの指示等が認識され、その認識結果が、成長モデル部31、感情／本能モデル部32、および行動決定部33に出力される。

#### 【0037】

即ち、センサ入力処理部30は、カメラ8が出力する画像データに基づいて、画像認識を行い、例えば、「ボールがある」とか、「壁がある」といったことを認識して、その認識結果を、成長モデル部31、感情／本能モデル部32、および行動決定部33に供給する。また、センサ入力処理部30は、マイク9が出力する音声データに基づいて、音声認識を行い、その音声認識結果を、成長モデル部31、感情／本能モデル部32、および行動決定部33に供給する。さらに、センサ入力処理部30は、圧力センサ10から与えられる圧力検出データを処理し、例えば、所定の閾値以上で、かつ短時間の圧力を検出したときには「たたかれた（しかられた）」と認識し、また、所定の閾値未満で、かつ長時間の圧力を検出したときには「なでられた（ほめられた）」と認識して、その認識結果を、成長モデル部31、感情／本能モデル部32、および行動決定部33に供給する。

## 【0038】

成長モデル部31は、ロボットの成長状態を規定するための成長モデルを管理している。成長モデルは、例えば、図5（A）に示すようなオートマトンで構成される。このオートマトンでは、成長状態は、ノード（状態）NODE<sub>0</sub>乃至NODE<sub>G</sub>で表現され、成長、つまり成長状態の遷移は、ある成長状態に対応するノードNODE<sub>g</sub>から、次の成長状態に対応するノードNODE<sub>g+1</sub>への遷移を表すアーケARC<sub>g+1</sub>で表現される（g=0, 1, …, G-1）。

## 【0039】

ここで、本実施の形態では、成長の状態は、図5（A）（後述する図5（B）においても同様）において、左のノードから右方向のノードに遷移していくようになっている。従って、図5（A）において、例えば、最も左のノードNODE<sub>0</sub>は、生まれたばかりの「新生児」の状態を表し、左から2番目のノードNODE<sub>1</sub>は、「幼児」の状態を表し、左から3番目のノードNODE<sub>2</sub>は「児童」の状態を表す。以下、同様にして、右方向のノードほど、より成長した状態を表し、最も右のノードNODE<sub>G</sub>は、「高齢」の状態を表す。

## 【0040】

そして、あるノードNODE<sub>g</sub>から、右隣のノードNODE<sub>g+1</sub>への遷移を表すアーケARC<sub>g+1</sub>には、その遷移が生じるための条件（入力）P<sub>tg+1</sub>が設定されており、ノードの遷移（成長）は、この条件に基づいて決定される。即ち、アーケARC<sub>g+1</sub>においては、その遷移が生じるために要求される、カメラ8や、マイク9、圧力センサ10からの出力、時間経過等についての条件P<sub>tg+1</sub>が設定されており、その条件P<sub>tg+1</sub>が満たされた場合に、ノードNODE<sub>g</sub>から右隣のノードNODE<sub>g+1</sub>への遷移が生じ、ロボットが成長する。

## 【0041】

なお、図5（A）に示した成長モデルにおいては、左のノードから右方向のノードに遷移していくパスが1つで固定であるため、成長過程は、その1つのパスにしたがった一定のものとなるが、成長モデルとしては、成長過程が複数パターン存在する、例えば、図5（B）に示すようなものを用いるようにすることも可能である。

## 【0042】

即ち、図5（B）においては、成長モデルは、木構造になっており、例えば、「新生児」に対応するノードNODE<sub>0</sub>からは、「幼児」に対応するノードNODE<sub>1-0</sub>またはNODE<sub>1-1</sub>のうちのいずれかに遷移し得る。「幼児」に対応する2つのノードNODE<sub>1-0</sub>およびNODE<sub>1-1</sub>は、例えば、人間で言えば、異なる人格の幼児を表しており、ノードNODE<sub>1-0</sub>またはNODE<sub>1-1</sub>のうちのいずれに遷移するかは、カメラ8や、マイク9、圧力センサ10からの出力、時間経過等に基づいて決定される。従って、この場合、受けた刺激等によって、ロボットの成長過程を異なるものとすることができます。

## 【0043】

なお、成長モデルとしては、木構造の他、グラフ構造等のオートマトンで表現可能な構造を採用することができる。

## 【0044】

成長モデル部31は、以上のような成長モデルに基づいて、ロボットの成長状態を管理し、現在の成長状態を、センサ入力処理部30および行動決定部33に出力する。

## 【0045】

一方、感情／本能モデル部32は、ロボットの、例えば、「うれしさ」や、「悲しさ」、「怒り」等の、いわゆる喜怒哀楽の感情の状態を表現する感情モデルと、例えば、「食欲」や、「睡眠欲」、「運動欲」等の本能の状態を表現する本能モデルとを管理している。

## 【0046】

即ち、感情モデルと本能モデルは、それぞれ、ロボットの感情と本能の状態を、例えば、0乃至100の範囲の整数値で表すもので、感情／本能モデル部32は、感情モデルや本能モデルの値を、センサ入力処理部30の出力や、行動決定部33の出力、時間経過に基づいて変化させる。従って、ここでは、ロボットの感情および本能の状態は、センサ入力処理部30が outputする外部の認識結果の他、行動決定部33の出力や、時間経過にもに基づいて変化する。そして、感情／本能モデル部32は、変更後の感情モデルおよび本能モデルの値（ロボットの感情

および本能の状態)を、行動決定部33に供給する。

#### 【0047】

なお、感情モデルや本能モデルは、その他、例えば、オートマトン等で構成することも可能である。

#### 【0048】

ここで、ロボットの感情および本能の状態は、行動決定部33の出力に基づいて、例えば、次のように変化する。

#### 【0049】

即ち、行動決定部33から感情／本能モデル部32に対しては、ロボットがとっている現在の行動または過去にとった行動の内容（例えば、「そっぽを向いた」など）を示す行動情報が供給されるようになっている。

#### 【0050】

いま、ロボットにおいて、「怒り」の感情が高いときに、使用者から何らかの刺激を受け、ロボットが、「そっぽを向く」行動をとったとする。この場合、行動決定部33から感情／本能モデル部32に対して、行動情報「そっぽを向いた」が供給される。

#### 【0051】

一般に、怒っているときに、そっぽを向くような不満を表現する行動をとると、その行動によって、怒りが幾分か静まることがある。そこで、感情／本能モデル部32は、行動決定部33から、行動情報「そっぽを向いた」が供給された場合、「怒り」を表す感情モデルの値を、小さい値に変更する（「怒り」の度合いが小さくなるように変更する）。

#### 【0052】

行動決定部33は、センサ入力処理部30が出力する外部の認識結果や、成長モデル部31における成長モデル、感情／本能モデル部32における感情モデルおよび本能モデル、行動モデル記憶部33Aの記憶内容、時間経過等に基づいて、次にとる行動を決定し、その行動を表す行動情報を、感情／本能モデル部32および姿勢遷移部34に供給する。

#### 【0053】

ここで、行動モデル記憶部33Aは、ロボットの行動を規定する、例えば、図6に示すような行動モデルを記憶している。

#### 【0054】

即ち、本実施の形態では、行動モデルは、例えば、確率オートマトンで構成される。この確率オートマトンでは、行動は、ノード（状態）NODE<sub>0</sub>乃至NODE<sub>M</sub>で表現され、行動の遷移は、ある行動に対応するノードNODE<sub>m0</sub>から、他の行動（同一の行動である場合もある）に対応するノードNODE<sub>m1</sub>への遷移を表すアーケARC<sub>m1</sub>で表現される（m0, m1 = 0, 1, …, M）。

#### 【0055】

なお、図6においては、図が煩雑になるのを避けるため、M+1個のノードからなる確率オートマトンに関して、ノードNODE<sub>0</sub>から、他のノード（自身を含む）NODE<sub>0</sub>乃至NODE<sub>M</sub>へのアーケARC<sub>0</sub>乃至ARC<sub>M</sub>のみを示してある。

#### 【0056】

あるノードNODE<sub>m0</sub>から、他のノードNODE<sub>m1</sub>への遷移を表すアーケARC<sub>m1</sub>には、その遷移が生じる遷移確率P<sub>t<sub>m1</sub></sub>が設定されており、また、各ノードNODE<sub>m</sub>（m=0, 1, …, M）には、そのノードNODE<sub>m</sub>に対応する行動が生起する生起確率P<sub>nm</sub>が設定されている。そして、行動モデルでは、ある行動の次に、どのような行動をとるかは、アーケに設定されている遷移確率と、ノードに設定されている生起確率とから決定される。

#### 【0057】

行動決定部33は、行動モデルの、現在の行動（以下、適宜、現在行動という）に対応するノードから、どのノードに遷移するかを、現在行動に対応するノードから延びるアーケに設定されている遷移確率、およびそのアーケによって示される遷移先のノードに設定されている生起確率の他、成長モデル部31における成長モデルや、感情／本能モデル部32における感情モデルおよび本能モデル、時間経過、センサ入力処理部30が出力する外部の認識結果に基づいて決定し、遷移後のノードに対応する行動（以下、適宜、遷移行動という）を表す行動情報と、感情／本能モデル部32および姿勢遷移部34に供給する。

## 【0058】

従って、行動決定部33では、例えば、センサ入力処理部30が出力する外部の認識結果が同一のものであっても、成長モデルが表す成長の状態によっては、異なる行動を起こすことが決定される。

## 【0059】

ここで、行動決定部33は、成長モデルに基づいて、行動モデルの各ノードに設定された生起確率を変更するようになっている。

## 【0060】

具体的には、例えば、いま、行動モデルにおける各アーケに設定された遷移確率を無視して、各ノードに設定された生起確率だけを考えると、各ノードに設定された生起確率の総和は、1となる。例えば、いま、行動モデルにおいて、行動として、「寝る」、「立ち上がる」、「歩く」、および「走る」の4つが存在するとして、ある行動の生起確率を、 $P$ （行動）で表すと、 $P$ （「寝る」）+ $P$ （「立ち上がる」）+ $P$ （「歩く」）+ $P$ （「走る」）は、1となる。

## 【0061】

ロボットの成長によっては、各ノードに設定された生起確率の総和が1であることは変わらないが、行動決定部33では、各ノードに対応する行動の生起確率が、ロボットの成長にしたがって変更される。

## 【0062】

即ち、例えば、成長モデルにおいて、成長状態が「新生児」を表している場合においては、行動決定部33は、行動「寝る」の生起確率を1にし、他の行動「立ち上がる」、「歩く」、および「走る」の生起確率を0にする。この場合、ロボットは、「寝る」以外の行動はしない。

## 【0063】

その後、時間経過等に基づいて、成長モデルにおいて、成長状態が「幼児」になると、行動決定部33は、行動「寝る」の生起確率を幾分小さくするとともに、行動「立ち上がる」の生起確率を、その分高くする。さらに、行動決定部33は、ロボットが成長するにつれて、行動「歩く」、「走る」の順に、それぞれの生起確率を徐々に高くしていく。これにより、ロボットは、最初は、「寝る」こ

としかしないが、成長するにつれて、「立ち上がる」ようになり、さらに、「歩く」、「走る」といった行動をとるようになる。即ち、起こし得る行動パターンが増加していく。

#### 【0064】

行動決定部33では、以上のようにして、センサ入力処理部30が出力する外部の認識結果が同一のものであっても、成長状態によって異なる行動を起こすことが決定される他、感情モデルや本能モデルの値によっても、異なる行動を起こすことが決定される。

#### 【0065】

具体的には、例えば、センサ入力処理部30の出力が、「目の前に手のひらが差し出された」旨を示すものである場合において、「怒り」の感情モデルが「怒っていない」ことを表しており、かつ「食欲」の本能モデルが「お腹がすいていない」ことを表しているときには、行動決定部33は、目の前に手のひらが差し出されたことに応じて、遷移行動として「お手」を行わせることを決定する。

#### 【0066】

また、例えば、上述の場合と同様に、センサ入力処理部30の出力が、「目の前に手のひらが差し出された」旨を示すものであり、かつ「怒り」の感情モデルが「怒っていない」ことを表しているが、「食欲」の本能モデルが「お腹がすいている」ことを表している場合には、行動決定部33は、遷移行動として、「手のひらをぺろぺろなめる」ような行動を行わせることを決定する。

#### 【0067】

さらに、センサ入力処理部30の出力が、「目の前に手のひらが差し出された」旨を示すものであるが、「怒り」の感情モデルが「怒っている」ことを表している場合には、「食欲」の本能モデルの値にかかわらず、行動決定部33は、遷移行動として、「ぷいと横を向く」ような行動を行わせることを決定する。

#### 【0068】

なお、行動決定部33は、所定のトリガ(trigger)があった場合に、行動モデルのノードを遷移し、遷移行動を決定する。即ち、行動決定部33は、例えば、ロボットが現在行動を開始してから所定の時間が経過した場合や、音声認識結果

等の特定の認識結果がセンサ入力処理部30から出力された場合、成長モデル部31における成長モデルが、ある成長状態を表している場合、感情／本能モデル部32における感情モデルや本能モデルの値が所定の閾値に達した場合等に、遷移行動を決定する。

#### 【0069】

また、行動決定部33は、ロータリエンコーダ $12_1$ 乃至 $12_N$ の出力に基づいて、現在のロボットの姿勢を認識し、その姿勢から、自然な形で遷移行動を起こすことができるような行動情報を、姿勢遷移部34に出力するようになっている。

#### 【0070】

姿勢遷移部34は、行動決定部33から供給される行動情報に基づいて、現在の姿勢から次の姿勢に遷移するための姿勢遷移情報を生成し、モータ制御部35に出力する。即ち、姿勢遷移部34は、現在の姿勢を、行動決定部33を介して、ロータリエンコーダ $12_1$ 乃至 $12_N$ の出力に基づいて認識し、行動決定部33からの行動情報に対応する行動（遷移行動）を、ロボットにとらせるためのモータ $7_1$ 乃至 $7_N$ の回転角や回転速度等を計算し、姿勢遷移情報として、モータ制御部35に出力する。

#### 【0071】

モータ制御部35は、姿勢遷移部34からの姿勢遷移情報にしたがって、モータ $7_1$ 乃至 $7_N$ を駆動するための駆動信号を生成し、モータ $7_1$ 乃至 $7_N$ に供給する。これにより、モータ $7_1$ 乃至 $7_N$ は回転駆動し、ロボットは、遷移行動を行う。

#### 【0072】

次に、図7は、図4のセンサ入力処理部30のうち、マイク9からの音声データを用いて音声認識を行う部分（以下、適宜、音声認識装置という）の機能的構成例を示している。

#### 【0073】

この音声認識装置においては、マイク9に入力された音声が、例えば、連続分布HMM(Hidden Markov Model)法にしたがって認識され、その音声認識結果が出力されるようになっている。

## 【0074】

即ち、特徴パラメータ抽出部41には、マイク9からの音声データが供給されるようになっており、特徴パラメータ抽出部41は、そこに入力される音声データについて、適当なフレームごとに、例えば、MFCC (Mel Frequency Cepstrum Coefficient)分析を行い、その分析結果を、特徴パラメータ（特徴ベクトル）として、マッチング部42に出力する。なお、特徴パラメータ抽出部41では、その他、例えば、線形予測係数、ケプストラム係数、線スペクトル対、所定の周波数帯域ごとのパワー（フィルタバンクの出力）等を、特徴パラメータとして抽出することが可能である。

## 【0075】

マッチング部42は、特徴パラメータ抽出部41からの特徴パラメータを用いて、音響モデル記憶部43、辞書記憶部44、および文法記憶部45を必要に応じて参照しながら、マイク9に入力された音声（入力音声）を、連続分布HMM法に基づいて音声認識する。

## 【0076】

即ち、音響モデル記憶部43は、音声認識する音声の言語における個々の音素や音節などの音響的な特徴を表す音響モデルを記憶している。ここでは、連続分布HMM法に基づいて音声認識を行うので、音響モデルとしては、HMM (Hidden Markov Model) が用いられる。辞書記憶部44は、認識対象の各単語について、その発音に関する情報（音韻情報）が記述された単語辞書を記憶している。文法記憶部45は、辞書記憶部35の単語辞書に登録されている各単語が、どのように連鎖する（つながる）かを記述した文法规則を記憶している。ここで、文法规則としては、例えば、文脈自由文法 (CFG) や、統計的な単語連鎖確率 (N-gram) などに基づく規則を用いることができる。

## 【0077】

マッチング部42は、辞書記憶部44の単語辞書を参照することにより、音響モデル記憶部43に記憶されている音響モデルを接続することで、単語の音響モデル（単語モデル）を構成する。さらに、マッチング部42は、幾つかの単語モデルを、文法記憶部45に記憶された文法规則を参照することにより接続し、そ

のようにして接続された単語モデルを用いて、特徴パラメータに基づき、連続分布HMM法によって、マイク9に入力された音声を認識する。即ち、マッチング部42は、特徴パラメータ抽出部41が出力する時系列の特徴パラメータが観測されるスコア（尤度）が最も高い単語モデルの系列を検出し、その単語モデルの系列に対応する単語列を、音声の認識結果として出力する。

#### 【0078】

具体的には、マッチング部42は、接続された単語モデルに対応する単語列について、時系列に入力される特徴パラメータの出現確率を累積し、その累積値をスコアとして、そのスコアを最も高くする単語列を、音声認識結果として出力する。マッチング部42による音声認識結果は、センサ入力処理部30の出力として、上述したように、成長モデル部31、感情／本能モデル部32、および行動決定部33に出力される。

#### 【0079】

ここで、辞書記憶部44における単語辞書に登録されているすべての単語にとってとり得る単語列を対象として、スコアを計算すると、その計算量が莫大になることから、どのような単語列を対象として、スコアの計算を行うかは、文法記憶部45に記憶された文法規則により制限される。また、マッチング部42では、出現頻度の高い単語列についてのスコアを高くするような重みが付される場合もある。

#### 【0080】

マッチング部42におけるスコア計算は、例えば、音響モデル記憶部43に記憶された音響モデルによって与えられる音響的なスコア（以下、適宜、音響スコアという）と、文法記憶部45に記憶された文法規則によって与えられる言語的なスコア（以下、適宜、言語スコアという）とを総合評価することで行われる。

#### 【0081】

即ち、あるN個の単語からなる単語列におけるk番目の単語を $W_k$ として、その単語 $W_k$ の音響スコアをA( $W_k$ )と、言語スコアをL( $W_k$ )と、それぞれ表すとき、その単語列のスコアSは、例えば、次式にしたがって計算される。

#### 【0082】

$$S = \sum (A(W_k) + C_k \times L(W_k)) \quad \dots \quad (1)$$

但し、 $\Sigma$ は、 $k$ を1からNに変えてのサメーションをとることを表す。また、 $C_k$ は、単語 $W_k$ の言語スコア $L(W_k)$ にかける重みを表す。

#### 【0083】

なお、音声認識装置が有する辞書記憶部44に記憶された単語辞書は、ロボットが扱うことのできる語彙を規定するから、その点で、非常に重要である。即ち、単語辞書に登録しておく単語数が多いほど、ロボットは、各種の発話に対して対処することができる。

#### 【0084】

但し、図7の音声認識装置は、図3のCPU20がプログラムを実行することにより機能的に実現されるものであり、この場合、CPU20の処理能力や、CPU20が利用可能なRAM22の空き領域の大きさ等によって、単語辞書における語彙数は制限を受ける。

#### 【0085】

また、音声波形は、例えば、同一の語彙が同一人により発話されたとしても、完全に一致することは、ほとんどなく、このため、マッチング部42におけるマッチング処理（スコア計算）は、曖昧性を有する入力に対する処理となることから、発話されていない単語列が音声認識結果とされる誤認識が生じる場合があるが、このような誤認識は、一般に、単語辞書に登録された語彙数が多くなると、類似する単語も多くなることに起因して増加する。

#### 【0086】

従って、単語辞書に登録する語彙は、CPU20の処理能力やRAM22の容量の他、要求される認識精度によっても制限を受ける。

#### 【0087】

成長スコア設定部46は、成長モデル部31における成長モデルに基づいて、音声認識装置による音声認識処理を制御する。即ち、図7の実施の形態では、成長スコア設定部46は、成長モデルに基づいて、辞書記憶部44における単語辞書を操作し、これにより、音声認識処理の制御を行うようになっている。

## 【0088】

具体的には、例えば、成長スコア設定部46は、成長モデルに基づき、単語辞書に登録された各単語について、ロボットの成長状態において理解することができる度合いを表す成長スコアを計算する。例えば、いま、ある単語Wを、ロボットの成長状態において理解することができる確率をP(W)と表すと、成長スコア設定部46は、単語Wに対する成長スコアG(W)を、例えば、次式にしたがって計算する。

## 【0089】

$$G(W) = -\log(P(W))$$

... (2)

## 【0090】

さらに、成長スコア設定部46は、成長スコアを、辞書記憶部44における単語辞書の、対応する単語に対応付ける。従って、単語辞書は、一般には、単語の表記等の単語を識別するための情報と、その単語の音韻情報とが登録されたものとなっているが、辞書記憶部44における単語辞書は、例えば、図8に示すように、単語の表記と音韻情報の他、その単語に対する成長スコアも登録されたものとなっている。ここで、図8においては、単語「走れ」、「起きろ」、「歩け」、「お手」、「お座り」、...に対して、10, 20, 30, 40, 50, ...の成長スコアがそれぞれ登録されている。

## 【0091】

辞書記憶部44における単語辞書に、図8に示したように、成長スコアが登録されている場合、マッチング部42では、単語列に対するスコアが、音韻スコアおよび言語スコアだけでなく、成長スコアも考慮して求められる。即ち、マッチング部42では、式(1)に替えて、例えば、式(3)にしたがって、スコアSが計算される。

## 【0092】

$$S = \sum (A(W_k) + C_k \times L(W_k) + D_k \times G(W_k))$$

... (3)

但し、 $G_k(W_k)$ は、k番目の単語 $W_k$ に対する成長スコアを表し、 $D_k$ は、その

成長スコアG ( $W_k$ ) にかける重みを表す。

#### 【0093】

以上のように、マッチング部42におけるスコアSの計算にあたって、成長スコアを導入することにより、ロボットの成長状態に応じて、音声認識装置による音声の認識精度が変化される。

#### 【0094】

即ち、例えば、成長状態が「幼児」である場合には、図8の単語辞書において、単語「走れ」の成長スコアが、他の単語の成長スコアに比較して小さくされ、これにより、発話「走れ」の音声認識精度が低くされる。その結果、例えば、「走れ」と発話がされた場合には、その発話を正確に音声認識して、ロボットに「走る」という行動をとらせるのが一般的であるが、ここでは、いわば故意に、発話「走れ」を誤認識させて、ロボットが「走る」という行動をとらせないようにする（あるいは、「走る」以外の行動をとらせるようにする）。

#### 【0095】

そして、ロボットが成長するにしたがって、単語「走れ」の成長スコアが徐々に大きくされ、その結果、発話「走れ」が正確に音声認識されるようになり、ロボットは、その発話にしたがって、「走る」という行動をとるようになる。

#### 【0096】

ここで、例えば、人間を例にすれば、成長（あるいは、学習）するにつれて、理解することのできる語彙が増加することから、上述のようにして、成長スコアに基づいて、音声認識精度を、成長モデルにしたがって変化させることは、実際の人間の成長に合致することになる。

#### 【0097】

次に、図9のフローチャートを参照して、図7の音声認識装置の動作について、さらに説明する。

#### 【0098】

マイク9に音声が入力され、その音声に対応するデジタルの音声データが、センサ入力処理部30に供給されると、その音声データは、特徴パラメータ抽出部41に供給される。

## 【0099】

特徴パラメータ抽出部41では、ステップS1において、音声データが音響分析されることにより、その特徴パラメータが時系列に抽出され、マッチング部42に供給される。マッチング部42は、ステップS2において、特徴パラメータ抽出部41からの特徴パラメータを用いて、式(3)にしたがい、成長スコアを考慮したスコアの計算(マッチング)を行い、そのスコアに基づいて、マイク9に入力された音声の音声認識結果を決定する。そして、ステップS3に進み、マッチング部42は、その音声認識結果を、成長モデル部31、感情／本能モデル部32、および行動決定部33に出力し、処理を終了する。

## 【0100】

なお、上述の場合には、単語辞書に登録された単語の成長スコアを、成長モデルから得られるロボットの成長状態に応じて変化(適応)させることで、音声認識装置の音声認識精度を変えるようにしたが、音声認識精度は、その他、例えば、ロボットの成長状態に応じて、音響モデル記憶部43における音響モデルや、文法記憶部45に記憶された文法規則を規定するパラメータを変更すること等によって変えるようにすることも可能である。

## 【0101】

また、音声認識装置では、単語辞書に登録されている単語を対象として音声認識が行われるため、単語辞書に登録する単語を変更することで、音声認識装置が音声認識の対象とする単語を変えるようにすることが可能である。即ち、例えば、成長モデルから得られるロボットの成長状態に応じて、順次、単語辞書に単語を登録したり、あるいは、単語辞書に登録された単語を削除し、音声認識の対象とする単語を変えるようにすることが可能である。

## 【0102】

さらに、上述の場合には、ロボットの成長状態に応じて、音声認識装置の音声認識精度を変化させるようにしたが、音声認識精度は、ロボットの感情や本能の状態に応じて変化させることも可能である。音声認識精度を、ロボットの感情や本能の状態に応じて変化させる場合には、図4に点線で示すように、感情／本能モデル部32からセンサ入力処理部30に対して、感情モデルおよび本能モデル

を供給するようにするとともに、図7に点線で示すように、これを、成長スコア設定部46で受信し、感情モデルや本能モデルから、成長スコアと同様の感情スコアや本能スコアを求めて、単語辞書に登録すれば良い。この場合、マッチング部42において、感情スコアや本能スコアを考慮して、最終的なスコアを求めるようになると、音声認識精度を、ロボットの感情や本能の状態に応じて変化させることができる。

#### 【0103】

即ち、感情スコアや本能スコアを考慮する場合には、ロボットの機嫌が良い状態のときには、例えば、単語「お手」や「お座り」の音声認識精度を向上させ、これにより、発話「お手」や「お座り」を正しく音声認識させ、対応する行動を起こさせることができる。また、ロボットの機嫌が悪い状態のときには、例えば、単語「お手」や「お座り」の音声認識精度を低下させ、これにより、発話「お手」や「お座り」を誤認識させて、発話と関係ない行動を起こさせて、ロボットの不機嫌さを表現することができる。

#### 【0104】

なお、図7の音声認識装置には、人間の音声以外にも、例えば、拍手の音や、ロボットの各部を叩く音、楽器の音、ロボットの周辺で発生する環境音、音楽（曲）等の各種の音を音声として音声認識させることができある。即ち、例えば、連続分布HMM法によって音声認識が行われる場合には、音声認識対象とする音声（音）をHMMにモデル化しておけば良い。そして、人間の音声以外の音声（音）を対象として音声認識を行う場合にも、その音声認識精度を、ロボットの成長状態や、感情の状態、本能の状態に応じて変化させることができる。

#### 【0105】

次に、音声認識を行う場合には、音声認識の対象とする単語すべてを、1つの単語辞書に登録しておく他、音声認識の対象とする単語を、複数のグループに分け（但し、異なるグループに、同一の単語が属していてもかまわない）、各グループに属する単語ごとに、単語辞書を構成しておくことがある。このように、音声認識対象とする単語を複数のグループに分けて、複数の単語辞書を用意する場合には、ある発話に対する音声認識処理は、一般に、その複数の単語辞書の全部

ではなく、そのうちの1つ、あるいは2以上の単語辞書を用いて行われる。

#### 【0106】

即ち、ある発話に対して、音声認識の対象とする語彙数は、上述のように、CPU20の処理能力やRAM22の容量の他、要求される認識精度によっても制限を受ける。

#### 【0107】

そこで、複数の単語辞書を用意しておき、ある発話の音声認識が、要求される認識精度等に基づいて、用いる単語辞書を選択して行われる場合がある。

#### 【0108】

また、音声認識の対象とする単語を、例えば、旅行、野球、ニュースなどといったドメイン（テーマ）ごとにグループ化して、複数の単語辞書を構成し、音声認識が行われる場面に応じて、用いる単語辞書を選択するようにして、特定のテーマに含まれる固有名詞（例えば、地名や、野球選手の名前、会社名など）などの音声認識性能を向上させる場合もある。

#### 【0109】

このように、複数の単語辞書を構成して、用いる単語辞書を選択し、音声認識を行うことは、図7の音声認識装置にも適用することができる。

#### 【0110】

即ち、例えば、いま、音声認識の対象とする単語を、Q個のグループに分け、これにより、各グループに対応するQ個の単語辞書 $D_1, D_2, \dots, D_Q$ を構成して、辞書記憶部44に記憶しておく。

#### 【0111】

さらに、成長スコア設定部46には、上述のQ個の単語辞書 $D_1$ 乃至 $D_Q$ それぞれについて、ロボットの成長状態に応じて割り当てる成長スコアを対応付けた、例えば、図10に示すような辞書成長スコアテーブルを作成して記憶しておく。

#### 【0112】

ここで、図10においては、例えば、図5（A）に示した成長モデルの各ノードで表されるロボットの成長状態に対して、各単語辞書 $D_1$ 乃至 $D_Q$ それぞれに対

する成長スコアが設定されている。

#### 【0113】

図10の辞書成長スコアテーブルによれば、現在のロボットの成長状態（を表すノード）から、単語辞書D<sub>1</sub>乃至D<sub>Q</sub>それぞれに対する成長スコアが決定される。

#### 【0114】

成長スコア設定部46は、単語辞書D<sub>1</sub>乃至D<sub>Q</sub>それぞれの単語に対して、辞書成長スコアテーブルに基づいて決定される成長スコアを、上述した場合と同様に登録し、マッチング部42では、その成長スコアを考慮して、マッチング処理が行われる。

#### 【0115】

但し、ある1つの単語が、2以上の単語辞書に登録されている場合には、その単語には、その2以上の単語辞書に対する成長スコアのうち、例えば、最も大きい値が登録される。また、いまの場合、例えば、成長スコアが0の単語は、音声認識の対象とされない。即ち、ここでは、0より大きい値の成長スコアが設定されている単語辞書に登録されている単語だけを対象として、音声認識が行われる。

#### 【0116】

以上のように、単語辞書に対して、成長スコアを設定することによっても、ロボットの成長状態に応じて、音声認識精度を変化させることができる。

#### 【0117】

即ち、例えば、いま、図10の辞書成長スコアテーブルにおいて、成長状態が「新生児」である場合の、単語辞書D<sub>1</sub>に対する成長スコアを70とするとともに、単語辞書D<sub>2</sub>に対する成長スコアを30とし、他の単語辞書D<sub>3</sub>乃至D<sub>Q</sub>に対する成長スコアを0としておく。さらに、成長状態が「幼児」である場合の、単語辞書D<sub>1</sub>乃至D<sub>3</sub>に対する成長スコアを、それぞれ40, 10, 50とするとともに、他の単語辞書D<sub>4</sub>乃至D<sub>Q</sub>に対する成長スコアを0としておく。

#### 【0118】

この場合、成長状態が「新生児」のときには、単語辞書D<sub>1</sub>に登録された単語

と、単語辞書D<sub>2</sub>に登録された単語のみを対象に音声認識が行われる。従って、単語辞書D<sub>1</sub>とD<sub>2</sub>に登録されている単語が、例えば、「はいはい」、「パパ」、および「ママ」の3単語だけであれば、この3単語に対応する発話のみが、正確に音声認識される。

## 【0119】

なお、いまの場合、単語辞書D<sub>1</sub>に対する成長スコアを70とするとともに、単語辞書D<sub>2</sub>に対する成長スコアを30としており、従って、単語辞書D<sub>1</sub>の成長スコアの方が、単語辞書D<sub>2</sub>の成長スコアよりも大きいため、単語辞書D<sub>2</sub>に登録されている単語よりは、単語辞書D<sub>1</sub>に登録されている単語の方が、より正確に音声認識される。

## 【0120】

その後、ロボットが成長し、成長状態が「幼児」になると、単語辞書D<sub>1</sub>およびD<sub>2</sub>の他、単語辞書D<sub>3</sub>に登録された単語も対象として音声認識が行われる。従って、単語辞書D<sub>3</sub>に登録されている単語が、例えば、「お手」および「お座り」の2単語であれば、上述の「はいはい」、「パパ」、および「ママ」の3単語に加えて、「お手」や「お座り」に対応する発話も、正確に音声認識することができるようになる。

## 【0121】

なお、いまの場合、成長状態が「新生児」から「幼児」になると、単語辞書D<sub>1</sub>に対する成長スコアは70から40に、単語辞書D<sub>2</sub>に対する成長スコアは30から10に、それぞれ低下する。従って、単語辞書D<sub>1</sub>およびD<sub>2</sub>に登録された単語の認識精度は、成長状態が「新生児」から「幼児」になると低下する。さらに、成長状態が「幼児」の場合には、単語辞書D<sub>3</sub>に対する成長スコアは50であり、単語辞書D<sub>1</sub>およびD<sub>2</sub>に対する成長スコアのいずれよりも大きい。その結果、成長状態が「幼児」になると、上述のように、「はいはい」、「パパ」、および「ママ」の3単語に加えて、「お手」や「お座り」に対応する発話も音声認識することが可能となるが、「はいはい」、「パパ」、および「ママ」に対応する発話の音声認識精度は、「お手」および「お座り」に対する発話の音声認識精度よりも劣化する。即ち、成長するにつれて、ロボットは、「はいはい」、「パパ

」、および「ママ」といった幼児期に使用される単語を理解することができなくなっていく。

#### 【0122】

以上のように、複数の単語辞書を構成して、用いる単語辞書を選択し、音声認識を行うことは、ハードウェア上等の制約がある場合には、その制約の範囲内で音声認識を行うことが可能となるから便宜である。

#### 【0123】

なお、上述のように複数の単語辞書を用意して音声認識を行う場合には、各単語辞書に対して成長スコアを割り当てる他、上述の感情スコアや本能スコアを割り当てるようにすることも可能である。

#### 【0124】

また、成長モデルとして、図5（A）ではなく、図5（B）に示したような木構造状のものを用いる場合には、図10の辞書成長スコアテーブルには、図5（B）の各ノードに対応する成長状態に、単語辞書の成長スコアを割り当てることが可能である。この場合、図5（B）において、例えば、「幼児」の成長状態は、ノードNODE<sub>1-0</sub>とNODE<sub>1-1</sub>の2状態があるが、同じ「幼児」の成長状態であっても、ノードNODE<sub>1-0</sub>に対応する成長状態となった場合と、ノードNODE<sub>1-1</sub>に対応する成長状態となった場合とで、音声認識可能な単語や、音声認識精度を、異なるものとすることができます。

#### 【0125】

即ち、例えば、ノードNODE<sub>0</sub>に対応する成長状態「新生児」から、ノードNODE<sub>1-0</sub>に対応する成長状態「幼児」に成長した場合には、成長状態が「新生児」のときに音声認識可能であった発話「はいはい」、「パパ」、および「ママ」に加えて、「お手」や「お座り」に対応する発話も音声認識可能なようになることができる。これに対して、ノードNODE<sub>0</sub>に対応する成長状態「新生児」から、ノードNODE<sub>1-1</sub>に対応する成長状態「幼児」に成長した場合には、成長状態が「新生児」のときに音声認識可能であった発話「はいはい」、「パパ」、および「ママ」は音声認識不可にして、「お手」や「お座り」に対応する発話だけ音声認識可能なようになることができる。

**【0126】**

次に、図7の音声認識装置において、辞書記憶部44における単語辞書に登録されていない単語（未知語）に対応する発話については、正確な音声認識結果を得ることができない。

**【0127】**

そこで、未知語の正確な音声認識を可能とするために、単語辞書には、新たに、単語を登録するようにすることができる。単語辞書への単語の登録は、例えば、ロボットに、入力用のインターフェースを設けてキーボード等を接続し、そのキーボードを操作することにより、新たな単語の表記と読み（音韻）を与えること等によって行うことが可能である。

**【0128】**

単語辞書に登録して音声認識の対象とする単語の数は、上述したように、CPU20やRAM22、あるいは認識精度上の問題から制約されることがあるが、ここでは、そのような問題による制約を受けない範囲で、単語辞書に新たに登録する単語の数を制限するようにすることができる。

**【0129】**

即ち、単語辞書に新たに登録する単語の数は、ロボットの成長状態に応じて制限することができる。具体的には、例えば、成長状態が「新生児」の場合は、新たに登録可能な単語数は、数単語とし、ロボットが成長するにしたがって、数十単語、数百単語と増加させていくことができる。この場合、ロボットが理解可能な単語が、成長に伴って増加していくことになる。

**【0130】**

なお、単語辞書に新たに登録する単語の数は、その他、上述の感情スコアや本能スコアに基づいて変化させることが可能である。

**【0131】**

また、センサ入力処理部30では、音声認識の他、カメラ8からの画像の認識が行われるが、この画像認識においても、あらかじめ認識可能な色や顔等のパターンに加えて、後から、新たに画像認識対象とする色や顔のパターンを登録することが可能である。そして、この場合も、新たに登録可能な色や顔のパターン数

を、ロボットの成長状態や、感情または本能の状態応じて変化させることが可能である。

#### 【0132】

次に、図7の音声認識装置では、上述したように、スコアを最も高くする単語列が、音声認識結果として出力されるが、一般に、音声認識結果は、正確なことが望ましい。

#### 【0133】

しかしながら、ロボットに適用される図7の音声認識装置では、故意に誤認識結果を出力することで、ロボットとのやりとりを楽しくすることが可能となる。即ち、例えば、漫才においては、会話の相手が、発話者の発話した単語を、音響的に類似する単語に捉えて、いわゆる「ボケ」を演出し、その「ボケ」に対して、発話者が、いわゆる「ツッコミ」を行う場合があるが、音声認識装置において、故意に、誤った音声認識結果を出力することで、このような「ボケ」や「ツッコミ」を、ロボットに行わせることが可能となる。

#### 【0134】

「ボケ」は、例えば、次のようにして実現することができる。

#### 【0135】

即ち、単語辞書には、図11に示すように、音声認識対象とする単語の表記および音韻情報に、その単語と音響的に類似しているが意味が異なる単語、あるいはその単語と音響的には類似していないが、その単語から連想される単語（以下、適宜、置換単語情報）を対応付けて登録しておく。そして、マッチング部42には、音声認識結果として得た、単語辞書に登録された単語ではなく、その単語に対応付けられている置換単語情報を、最終的な音声認識結果として出力させる。

#### 【0136】

この場合、行動決定部33では、その最終的な音声認識結果である置換単語情報に基づいて、次に起こす行動が決定されるため、ロボットは、使用者の発話内容に対応しない「ボケ」た行動を行うことになる。

#### 【0137】

なお、このような「ボケ」は、その他、例えば、上述のように、複数の単語辞書を用意しておき、ある単語辞書を用いるべきところを、他の単語辞書を用いて音声認識を行うことで、使用者の発話した単語と音響的に類似した他の単語を、音声認識結果として得ることができるようにすることによっても実現可能である。

#### 【0138】

ここで、上述のように、音声認識装置に、故意に、誤った音声認識結果を出力させることは、例えば、「ボケ」るか、「ボケ」ないかの2状態を有する状態遷移モデル等を用いて、ランダムに、しかも稀に行うようにするのが望ましい。また、この場合、状態遷移モデルにおける状態の遷移の仕方は、例えば、成長モデルや、感情モデル、本能モデルに基づいて変更することが可能である。

#### 【0139】

次に、使用者が「ボケ」た発話を行った場合には、ロボットにおいて、その「ボケ」に対して、「ツッコミ」の行動を行うようにすることが、例えば、上述の「ボケ」を実現する場合と同様に可能である。

#### 【0140】

即ち、音声認識装置において、使用者が「ボケ」た発話（誤った発話）を行ったかどうかを認識することは困難であるから、図11に示したように、音声認識対象とする単語の表記および音韻情報に、置換単語情報を対応付けて、単語辞書に登録しておき、マッチング部42には、音声認識結果として得た、単語辞書に登録された単語ではなく、その単語に対応付けられている置換単語情報を、最終的な音声認識結果として出力させる。

#### 【0141】

使用者が「ボケ」た発話を行った場合には、その発話に対応して、音声認識装置が出力する音声認識結果としての置換単語情報は、使用者が本来発話すべきであった単語となり、行動決定部33では、そのような置換単語情報に基づいて、次に起こす行動が決定されるため、ロボットは、使用者の「ボケ」た発話に対して、「ツッコミ」の行動を行うことになる。

#### 【0142】

以上のように、ロボットの状態に基づいて、音声認識処理を制御するようにしたので、ロボットの成長等の状態にあわせて変化する音声認識処理が行われ、その結果、エンタテイメント性の高いロボットを提供することなどが可能となる。

#### 【0143】

なお、本実施の形態においては、本発明を、エンターテイメント用のロボット（疑似ペットとしてのロボット）に適用した場合について説明したが、本発明は、これに限らず、例えば、産業用のロボット等の各種のロボットに広く適用することが可能である。

#### 【0144】

また、本実施の形態においては、ロボットに行動を行わせるための駆動手段としてモータ $7_1$ 乃至 $7_N$ を用いるようにしたが、本発明はこれに限らず、要は、外界に対して作用をもたらす行動（動作）を発現できるのであれば、駆動手段として、他のアクチュエータや、スピーカ、ブザー、照明装置等を広く用いることができる。

#### 【0145】

さらに、本実施の形態においては、上述した一連の処理を、CPU20にプログラムを実行させることにより行うようにしたが、一連の処理は、それ専用のハードウェアによって行うことも可能である。

#### 【0146】

なお、プログラムは、あらかじめプログラムメモリ21（図3）に記憶させておく他、フロッピーディスク、CD-ROM(Compact Disc Read Only Memory), MO(Magneto optical)ディスク、DVD(Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体に、一時的あるいは永続的に格納（記録）しておくことができる。そして、このようなリムーバブル記録媒体を、いわゆるパッケージソフトウェアとして提供し、ロボット（プログラムメモリ21）にインストールするようにすることができる。

#### 【0147】

また、プログラムは、リムーバブル記録媒体からインストールする他、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、無線で転送した

り、LAN(Local Area Network)、インターネットといったネットワークを介して、有線で転送し、プログラムメモリ21にインストールすることができる。

#### 【0148】

この場合、プログラムがバージョンアップされたとき等に、そのバージョンアップされたプログラムを、プログラムメモリ21に、容易にインストールすることができる。

#### 【0149】

ここで、本明細書において、CPU20に各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

#### 【0150】

また、プログラムは、1のCPUにより処理されるものであっても良いし、複数のCPUによって分散処理されるものであっても良い。

#### 【0151】

さらに、本実施の形態では、連続分布HMM法にしたがって、音声認識を行うようにしたが、音声認識は、連続分布HMM法による他、例えば、スペクトル波形どうしのマッチング等によって行うことも可能である。

#### 【0152】

##### 【発明の効果】

本発明の音声認識装置および音声認識方法、並びに記録媒体によれば、ロボットの状態に基づいて、音声認識処理が制御されるので、ロボットの成長等の状態にあわせて変化する音声認識処理を実現することができ、その結果、ロボットのエンタテイメント性の向上等を図ることが可能となる。

##### 【図面の簡単な説明】

###### 【図1】

本発明を適用したロボットの一実施の形態の外観構成例を示す斜視図である。

###### 【図2】

ロボットの内部構成を示すブロック図である。

【図3】

制御部11のハードウェア構成例を示すブロック図である。

【図4】

制御部11がプログラムが実行することにより実現される、制御部11の機能的構成例を示すブロック図である。

【図5】

成長モデルを示す図である。

【図6】

行動モデルを示す図である。

【図7】

センサ入力処理部30における音声認識を行う部分としての音声認識装置の構成例を示すブロック図である。

【図8】

単語辞書を示す図である。

【図9】

図7の音声認識装置の動作を説明するためのフローチャートである。

【図10】

辞書成長スコアテーブルを示す図である。

【図11】

「ボケ」を実現するための単語辞書を示す図である。

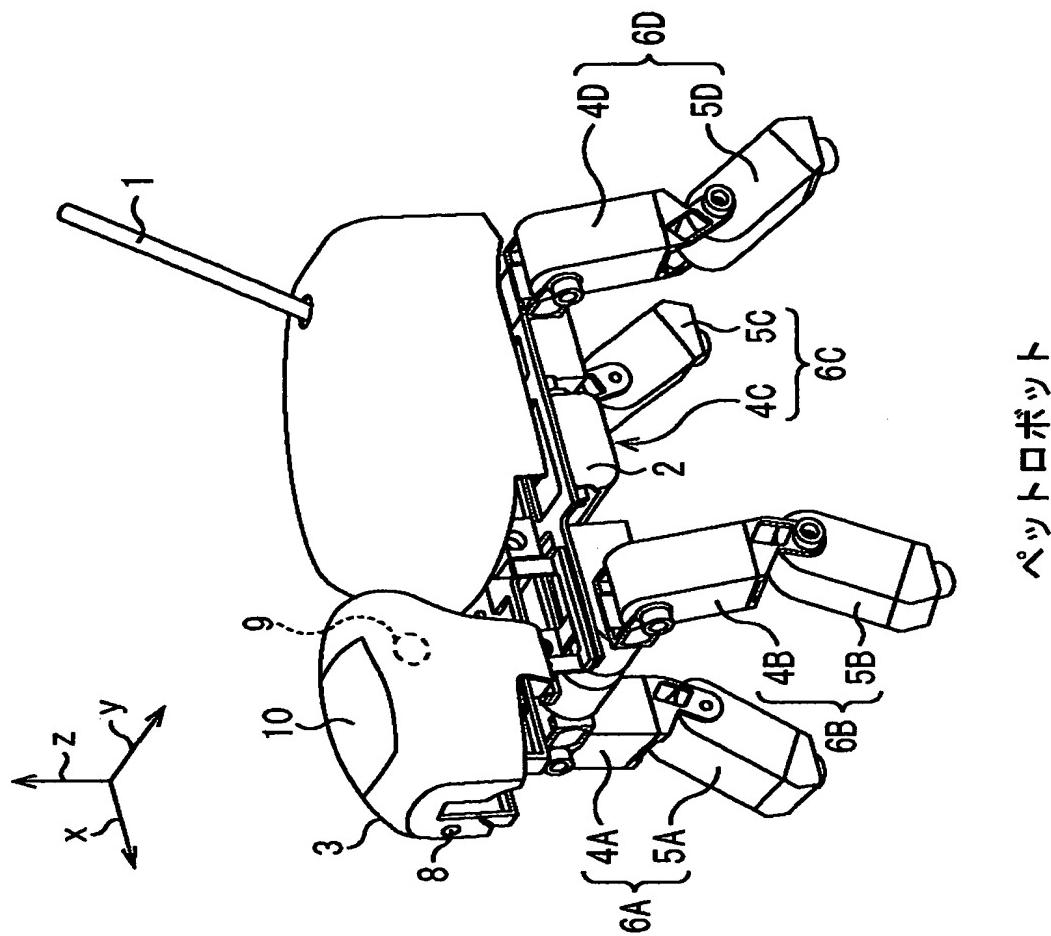
【符号の説明】

- 1 尻尾ユニット, 2 胴体部ユニット, 3 頭部ユニット, 4 A乃至  
4 D 大腿部ユニット, 5 A乃至5 D 脛部ユニット, 6 A乃至6 D 脚部  
ユニット, 7<sub>1</sub>乃至7<sub>N</sub> モータ, 8 カメラ, 9 マイク, 10 圧力  
センサ, 11 制御部, 12<sub>1</sub>乃至12<sub>N</sub> ロータリエンコーダ, 20 C  
PU, 21 プログラムメモリ, 22 RAM, 23 不揮発性メモリ,  
24 I/F, 25 モータドライバ, 30 センサ入力処理部, 31  
成長モデル部, 32 感情／本能モデル部, 33 行動決定部, 33A  
行動モデル記憶部, 34 姿勢遷移部, 35 モータ制御部, 41 特

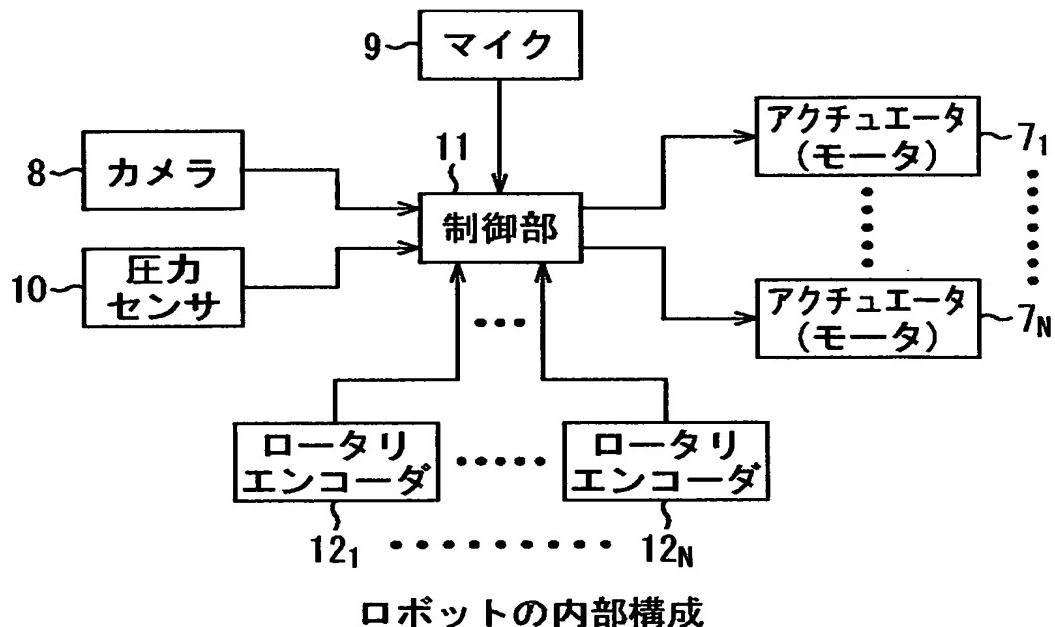
徴バラメータ抽出部, 42 マッチング部, 43 音響モデル記憶部, 4  
4 辞書記憶部, 45 文法記憶部, 46 成長スコア設定部

【書類名】 図面

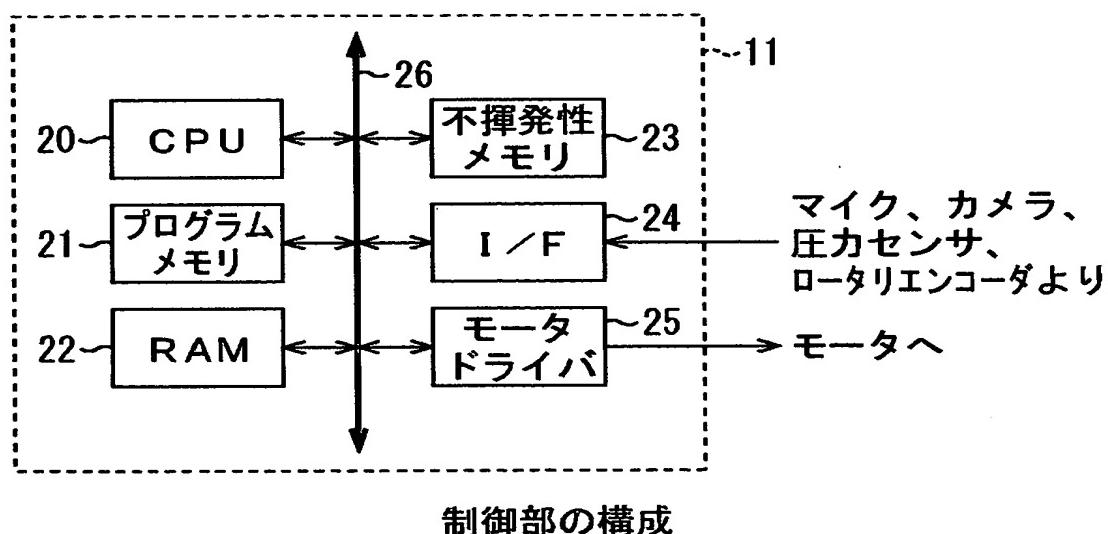
【図1】



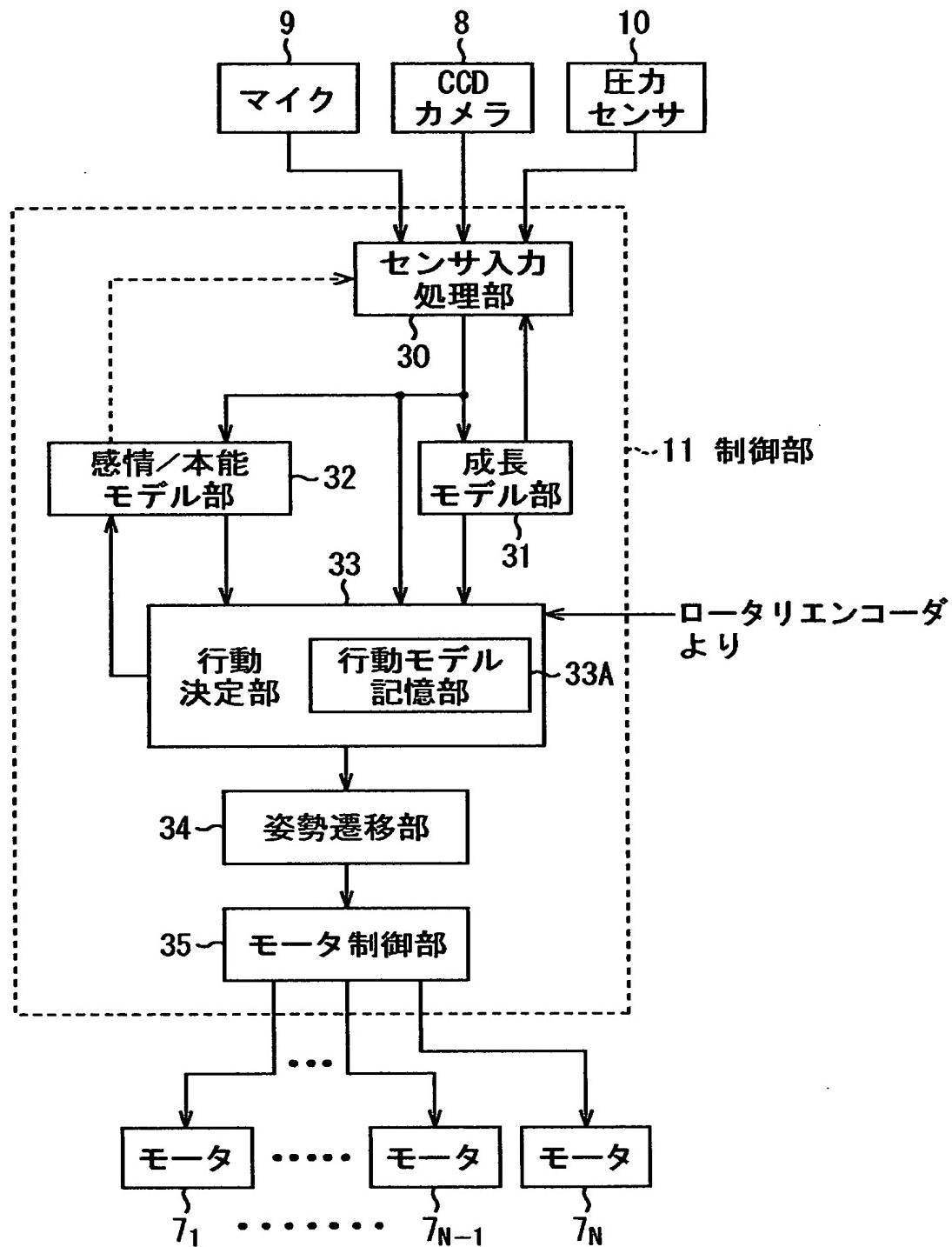
【図2】



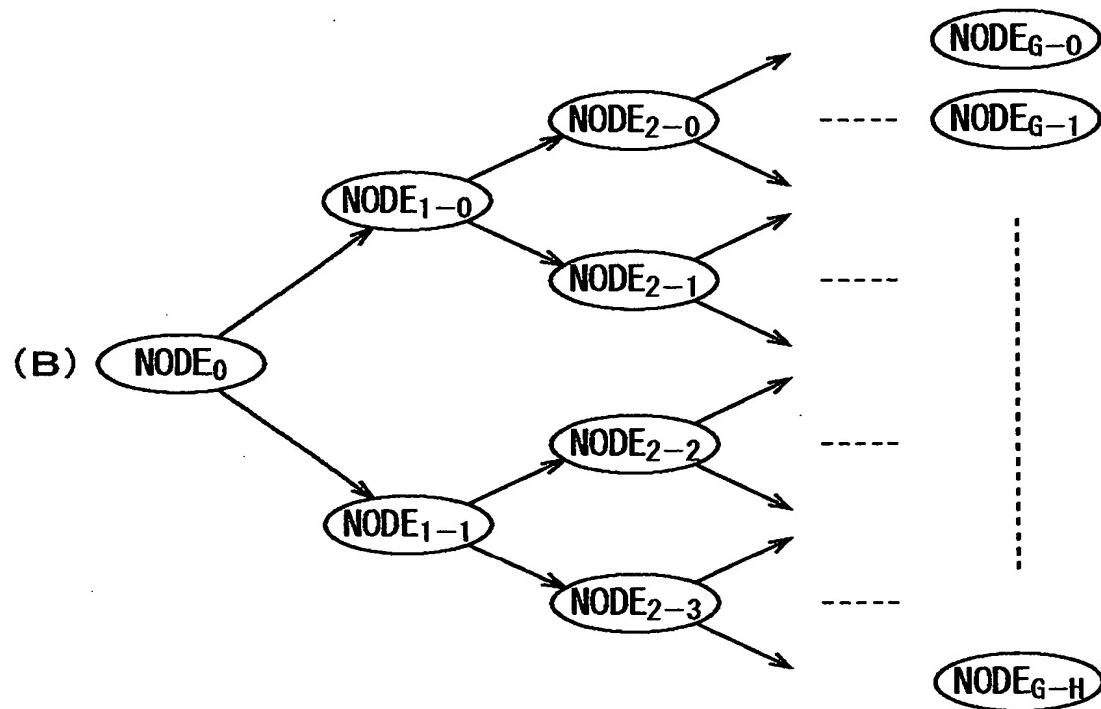
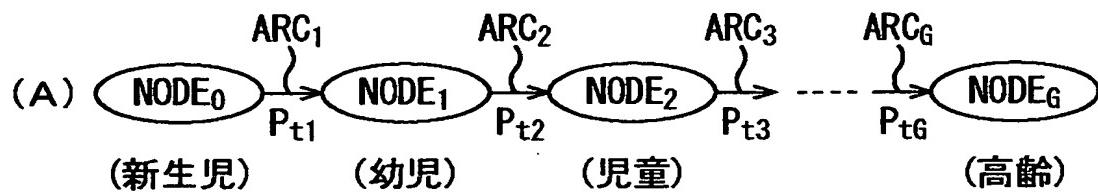
【図3】



【図4】

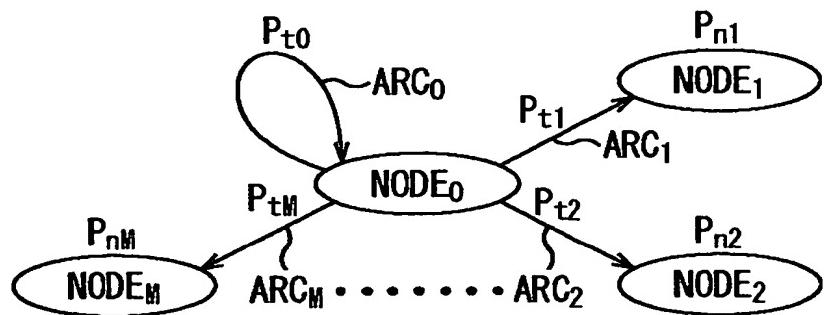


【図5】



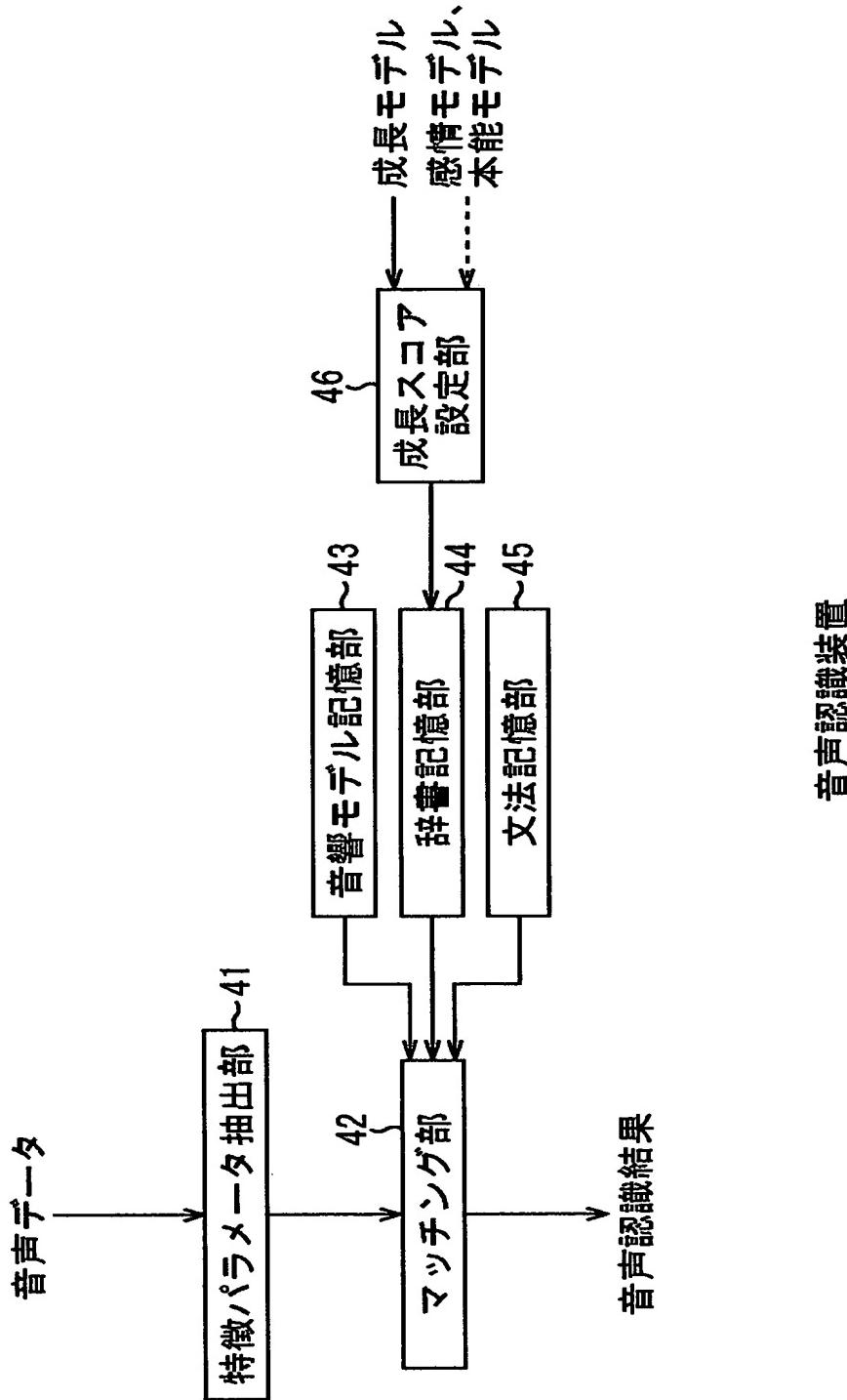
成長モデル(オートマトン)

【図6】



行動モデル(確率オートマトン)

【図7】

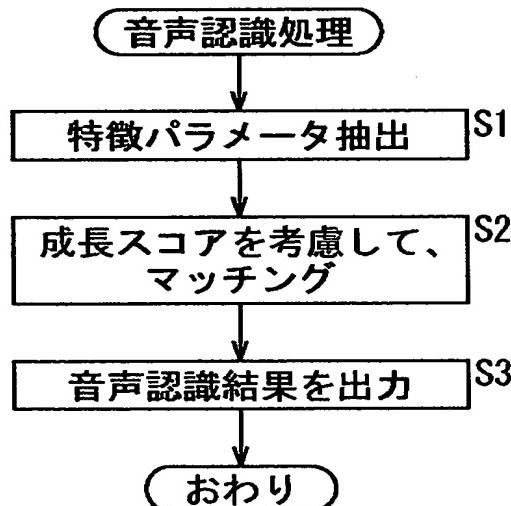


【図8】

表記	音韻情報	成長スコア
走れ	hashire	10
起きろ	okiro	20
歩け	aruke	30
お手	ote	40
お座り	osuwar i	50
⋮	⋮	⋮

単語辞書

【図9】



【図10】

成長状態	成長スコア			
	辞書D <sub>1</sub>	辞書D <sub>2</sub>	...	辞書D <sub>Q</sub>
NODE <sub>0</sub>				
NODE <sub>1</sub>				
⋮	⋮	⋮	⋮	⋮
NODE <sub>G</sub>				

辞書成長スコアテーブル

【図11】

表記	音韻情報	置換単語情報

単語辞書

【書類名】 要約書

【要約】

【課題】 ロボットの成長の状態にあわせて変化する音声認識処理を実現して、ロボットのエンタテイメント性の向上を図る

【解決手段】 成長スコア設定部46は、ロボットの成長状態に基づいて、辞書記憶部44における単語辞書に登録された各単語に、成長スコアを設定する。そして、マッチング部42は、単語辞書に登録された単語を対象とする音声認識を、その単語に設定された成長スコアを考慮して行う。

【選択図】 図7

出願人履歴情報

識別番号 [000002185]

1. 変更年月日 1990年 8月30日

[変更理由] 新規登録

住 所 東京都品川区北品川6丁目7番35号

氏 名 ソニー株式会社